DOCUMENT RESUME

ED 411 519 CS 215 966

AUTHOR McLean, James E.; Snyder, Scott W.; Abbott, Gypsy; Reid, Brian; Ernest, James; Heath, Roy

TITLE Evaluation of the Alabama Direct Assessment of Writing Program.

INSTITUTION SERVE: SouthEastern Regional Vision for Education.; Alabama Univ., Birmingham. School of Education.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

PUB DATE 1997-00-00

NOTE 36p.

CONTRACT RJ96006701

PUB TYPE Reports - Research (143) -- Tests/Questionnaires (160)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Grade 5; Grade 7; Intermediate Grades; Junior High Schools; Program Effectiveness; Psychometrics; State Standards; Teacher Attitudes; Test Results; *Writing Achievement; Writing Instruction; *Writing Tests

IDENTIFIERS Alabama; *Alabama Direct Assessment of Writing

ABSTRACT

A study evaluated the Alabama Direct Assessment of Writing Program. The comprehensive evaluation examined the background literature, the psychometric characteristics, and the impact of the writing assessment both quantitatively and qualitatively. Results indicated that: (1) the majority of fifth- and seventh-grade students in Alabama do not perform at acceptable standards in any of the major types of writing; (2) considerable gains in writing performance were evident in seventh grade between 1994 and 1995 across all types of writing; (3) moderate but sufficient agreement was evident between the two principal scorers of each writing response; (4) teachers were in considerable agreement that the state-mandated writing assessment positively influenced the teaching of writing; (5) higher-performing schools tended to place great value on writing and applied year-long systematic approaches to writing instruction; (6) in the lower-achieving schools, teachers did not appear to have the level of skill necessary to develop a cohesive writing curriculum effectively using the writing assessment results; and (7) the more positive that teachers' perceptions of the State Direct Assessment of Writing were, the higher the average student performance on the writing assessment at the school tended to be. The basic findings support the psychometric integrity of the writing assessment and its role in encouraging the teaching of writing in Alabama schools. (Contains 33 references and 10 tables of data; appendixes contain survey instruments.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 519

R&D

Evaluation of the Alabama Direct Assessment of Writing Program

CS215966

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

SERVE

Evaluation of the Alabama Direct Assessment of Writing Program

James E. McLean, Director
Scott W. Snyder, Associate Director

with assistance from
Gypsy Abbott
Brian Reid
James Ernest
Roy Heath

Center for Educational Accountability
School of Education
The University of Alabama at Birmingham
901 13th Street, South
Birmingham, AL 35294-1250
Phone: 205-934-7598
Fax: 205-975-5389

1997

funded by
SouthEastern Regional Vision for Education (SERVE)
associated with the
University of North Carolina at Greensboro
P. O. Box 5367
Greensboro, NC 27435

BEST COPY AVAILABLE

This document was produced with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract no. RJ96006701.

Edited by

Charles Ahearn, Director of Publications/Senior Editor
Stephen Chapman, Editor/Communications Specialist

Design

Christa Karantinos, Artist/Illustrator

The content of this publication does not necessarily reflect the views or policies of the Office of Educational Research and Improvement, U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Contents

About the SERVE Organization 4

Executive Summary 6

Introduction 7

Relevant Literature 8

Psychometric Analysis 12

Tables

1 Proportion of Students Scoring at Acceptable Levels or Above on Writing Performance 13

2 Percentages of Students Within Clusters Above the Acceptable Level on the 1995 Alabama Direct Writing Assessment 14

3 Summary of Discrepancy, Correlation, and Shared Variance by Year 15

Empirical Evaluation 16

Tables

4 Description of Interviewees 17

5 Number and (Percentages) Responding "Yes" to Interview Items 18

6 How the Rubric Was Used by Interviewees 19

7 Demographic and Involvement Variables from Survey 20

8 How the Rubric Was Used by Survey Respondents 20

9 Responses to Survey Items 21

10 Correlations of Predicted Writing Performance with Item Responses 22

Summary and Conclusions 24

References 25

Appendices 27

Ordering Information 31

About the SERVE Organization

SERVE, the SouthEastern Regional Vision for Education, is a consortium of educational organizations whose mission is to promote and support the continual improvement of educational opportunities for all learners in the Southeast. Formed by a coalition of business leaders, governors, policymakers, and educators seeking systemic, lasting improvement in education, the organization is governed and guided by a Board of Directors that includes the chief state school officers, governors, and legislative representatives from Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina. Committed to creating a shared vision of the future of education in the Southeast, the consortium impacts educational change by addressing critical educational issues in the region, acting as a catalyst for positive change and serving as a resource to individuals and groups striving for comprehensive school improvement.

SERVE's core component is a regional educational laboratory funded since 1990 by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. Building from this core, SERVE has developed a system of programs and initiatives that provides a spectrum of resources, services, and products for responding effectively to national, regional, state, and local needs. SERVE is a dynamic force, transforming national education reform strategies into progressive policies and viable initiatives at all levels. SERVE Laboratory programs and key activities are centered around

- Applying research and development related to improving teaching, learning, and organizational management
- Serving the educational needs of young children and their families more effectively
- Providing field and information services to promote and assist local implementation of research-based practices and programs
- Offering policy services, information, and assistance to decision makers concerned with developing progressive educational policy
- Connecting educators to a regional computerized communication system, so that they may search for and share information, and network
- Developing and disseminating publications and products designed to give educators practical information and the latest research on common issues and problems

The Eisenhower Consortium for Mathematics and Science Education at SERVE is part of the national infrastructure for the improvement of mathematics and science education sponsored by OERI. The consortium coordinates resources, disseminates exemplary instructional materials, and provides technical assistance for implementing teaching methods and assessment tools.

The SouthEast and Islands Regional Technology in Education Consortium (SEIR◆TEC) serves 14 states and territories. A seven-member partnership led by SERVE, the consortium offers a variety of services to foster the infusion of technology into K-12 classrooms. The Region IV Comprehensive Assistance Center provides a coordinated, comprehensive approach to technical assistance through its partnership with SERVE.

A set of special purpose institutes completes the system of SERVE resources. These institutes provide education stakeholders extended site-based access to high quality professional development programs, evaluation and assessment services, training and policy development to improve school safety, and subject area or project-specific planning and implementation assistance to support clients' school improvement goals.

Following the distributive approach to responding and providing services to its customers, SERVE has ten offices in the region. The North Carolina office at the University of North Carolina at Greensboro is headquarters for the Laboratory's executive services and operations. Policy offices are located in the departments of education in Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina.

SERVE-Alabama

Policy
Office forthcoming—please call any Serve Policy office for assistance

SERVE-Florida

Early Childhood, Publications, Lab, Field Services
345 South Magnolia Drive
Suite D-23
Tallahassee, FL 32301
904-671-6000
800-352-6001
Fax 904-671-6020

Database Information Services Clearinghouse
345 South Magnolia Drive
Suite E-21
Tallahassee, FL 32301
904-671-6012
800-352-3747
Fax 904-671-6010
SERVE-Line (modem only)
800-487-7605

Eisenhower Consortium for Mathematics and Science Education at SERVE
345 South Magnolia Drive
Suite E-22
Tallahassee, FL 32301
904-671-6033
800-854-0476
Fax 904-671-6010

Policy
345 South Magnolia Drive
Suite D-23
Tallahassee, FL 32301
904-671-6029
Fax 904-671-6020

Commissioner of Education's Office
The Capitol
LL 24
Tallahassee, FL 32399
904-488-9513
Fax 904-488-1492

SERVE-Georgia

Technology, Field Services
41 Marietta Street, NW
Suite 1000
Atlanta, GA 30303
404-893-0100
800-659-3204
Fax 404-577-7812
SERVE-Line (modem only)
800-487-7605

Policy
Georgia Department of Education
2054 Twin Towers East
Atlanta, GA 30334
404-657-0148
Fax 404-651-5231

SERVE-Mississippi

Delta Project
Delta State University
P.O. Box 3183
Cleveland, MS 38733
601-846-4384
800-326-4548
Fax 601-846-4402

Policy
Mississippi Department of Education
P.O. Box 771
Jackson, MS 39205
601-359-3501
Fax 601-359-3667

SERVE-North Carolina

Executive Services, Operations, Research and Development
P.O. Box 5367
Greensboro, NC 27435
910-334-3211
800-755-3277
Fax 910-334-3268

Policy
Department of Public Instruction
Education Building
301 North Wilmington Street
Raleigh, NC 27601-2825
919-715-1245
Fax 919-715-1278

SERVE-South Carolina

Policy
South Carolina Department of Education
1429 Senate Street
1005 Rutledge Building
Columbia, SC 29201
803-734-8496
Fax 803-734-3389

SERVE, Inc.-North Carolina

Business Office
P.O. Box 5406
Greensboro, NC 27435
910-334-4669
910-334-4670
800-545-7075
Fax 910-334-4671

SouthEast and Islands Regional Technology in Education Consortium
41 Marietta Street, NW
Suite 1000
Atlanta, GA 30303
404-893-0100
800-659-3204
Fax 404-577-7812
E-mail: seirtec@serve.org

Region IV Comprehensive Center
P.O. Box 5406
Greensboro, NC 27435
910-334-4667
800-545-7075
Fax 910-334-4671
E-mail: regivcc@serve.org

Evaluation and Assessment Services
P.O. Box 5367
Greensboro, NC 27435
910-334-3211
800-755-3277
Fax 910-334-3268

Professional Development Institute (PDI)
P.O. Box 5406
Greensboro, NC 27435
910-334-4667
800-545-7075
Fax 910-334-4671
E-mail: pdi@serve.org

Southeastern Regional Safe Schools Institute
P.O. Box 5406
Greensboro, NC 27435
910-334-4664
910-334-4665
800-545-7075
Fax 910-334-4671

Roy H. Forbes, Ed.D.
Executive Director

http//:www.serve.org

E-mail: info@serve.org

Executive Summary

The Center for Educational Evaluation at The University of Alabama at Birmingham was contracted to perform an evaluation of the Alabama Direct Assessment of Writing Program (hereafter known as the Assessment) for the Alabama State Department of Education. The evaluation was performed under the auspices and funding of the SouthEastern Regional Vision for Education (SERVE), a federally funded regional research and service laboratory. The comprehensive evaluation examined the background literature, the psychometric characteristics, and the impact of the writing assessment both quantitatively and qualitatively.

The evaluation of the Alabama Direct Assessment of Writing Program illuminated a number of key issues and outcomes. These findings extend the existing knowledge base about state-level performance assessments as well as provide evidence that should be useful to policy-makers in Alabama. Key conclusions are

- The majority of 5th and 7th grade students in Alabama do not perform at acceptable standards in any of the major types of writing.
- Considerable gains in writing performance were evident in 7th grade between 1994 and 1995 across all types of writing. This improvement appears to reflect improved educational opportunities for students rather than statistical or measurement artifacts.
- Moderate but sufficient agreement was evident between the two principal scorers of each writing response. While the interrater reliabilities would not be adequate to make high-stakes decisions about individual students, they are sufficiently strong to make aggregate judgments. The state's scoring system uses Gearhart, Herman, Novak, Wolf, and Abedi's (1994) recommendation for ensuring adequate reliability in that the scores for the independent raters are combined when determining performance.
- Teachers are in considerable agreement that the state-mandated writing assessment has positively influenced the teaching of writing and that the emphasis on student writing would diminish if the writing assessment were eliminated.
- High-performing schools tended to place great value on writing and applied year-long systematic approaches to writing instruction. The use of the scoring rubrics for instructional planning was among the strategies that appear to differentiate the high-performing and improving schools from the low-performing schools.
- In the lower-achieving schools, the teachers did not appear to have the level of skill necessary to develop a cohesive writing curriculum effectively using the writing assessment results. This suggests that they will likely need assistance to improve student writing skills.
- The more positive that teachers' perceptions of the State Direct Assessment of Writing are, the higher the average student performance on the writing assessment at the school tends to be.

The basic findings support the psychometric integrity of the writing assessment and its role in encouraging the teaching of writing in Alabama schools. The evaluation also found that for the Alabama Direct Assessment of Writing to be of greatest benefit, many teachers in the less-successful schools must be provided with additional support to learn how to deal more effectively with the information they gain. The writing assessment has a positive overall influence on the ability of students in Alabama to communicate effectively through the written word.

Introduction

The Center for Educational Evaluation at The University of Alabama at Birmingham was contracted to perform an evaluation of the Alabama Direct Assessment of Writing Program for the Alabama State Department of Education. The evaluation was performed under the auspices and funding of the SouthEastern Regional Vision for Education (SERVE), a federally funded regional research and service laboratory. The evaluation had the following objectives:

- Provide a framework based on the literature relevant to large-scale direct assessment of writing and the impact of testing programs on curriculum
- Perform an analysis of the psychometric properties of the Alabama Direct Assessment of Writing as it was implemented in grades 5 and 7 during the period 1992 through 1995
- Conduct an empirical evaluation of the Alabama Direct Assessment of Writing Program using both qualitative and quantitative methods with an emphasis on determining the perceived impact of the writing assessment on the teaching of writing in Alabama schools

The evaluation was conducted in May and June of 1996. The lateness in the school year posed some restrictions on the techniques that could be used. The primary limitation of this schedule was that many schools completed their years in mid-to-late May and teachers were no longer available. The methods and results of the evaluation are presented in three sections: Relevant Literature, Psychometric Analysis, and Empirical Evaluation. A Summary and Conclusions section is also included.

Relevant Literature

The review of the relevant literature covers four general areas—the Alabama Direct Assessment of Writing Program, national direct assessments of writing, validity of assessing writing directly, and the impact of assessments on curriculum. These areas are summarized and synthesized in a summary section.

The Alabama Direct Assessment of Writing Program

The Alabama Direct Assessments of Writing were developed to assess the writing skills of Alabama students directly (Alabama State Department of Education, undated d). That is, students' ability to write would be assessed based on the communication value of examples of their writing rather than on their knowledge of the rules of writing. It was originally intended to be given in grades 2, 5, and 7. However, it is now given only at grades 5 and 7. Philosophically, the Alabama Direct Assessment of Writing Program was developed to move writing instruction forward (Alabama State Department of Education, undated d). Consequently, the program assessed all modes of writing (i.e., descriptive, narrative, expository, and persuasive for Grade 7 only) to encourage that all modes of writing would be taught rather than only the one that was to be assessed that year (Alabama State Department of Education, undated a, b, & c; CTB Macmillan / McGraw-Hill, 1992).

In 1991, the Alabama Direct Assessment of Writing: Grade Five was piloted statewide by Measurement Incorporated of Durham, North Carolina (Alabama State Department of Education, undated d). One year later, in 1992, Measurement Incorporated conducted a sample pilot of the Alabama Direct Assessment of Writing: Grade Seven. Both assessments were scored according to Alabama's 6-point focused holistic rubric and, later, by Alabama's 4-point analytic rubric. The analytic rubric gave a score for each of the following: writing mechanics, sentence formation, and grammar and usage (Alabama State Department of Education, undated d). During the 1994-95 school year, the Alabama State Department of Education elected to report student achievement at four levels (Alabama State Department of Education, undated d). This change required the writing of new rubrics for both the Grade 5 and Grade 7 writing assessments. In order to make a smooth transition from the 6-point scale to the 4-point scale, it was necessary to rescore papers from the pilot assessments and prepare anchor sets, training sets, qualifying sets, and calibration sets for each prompt that would be used for the years 1995, 1996, and 1997 in Grade 5 and for the years 1995, 1996, 1997, and 1998 in Grade 7.

In February of 1995, the Psychological Corporation (Harcourt Brace Educational Measurement), the current contracting company for both Alabama Direct Assessments of Writing, rescored the student responses from the Grade 7 pilot. In March 1995, the company also rescored the responses from the Grade 5 pilot.

While there is not a direct mapping of the 6-point rubric onto the 4-point rubric, there is one point in common on the two rubrics. That point falls between scores 3 and 4 on the 6-point rubric and between scores 2 and 3 on the 4-point rubric. Thus, performance of students scored by the two rubrics can be compared by grouping scores 1, 2, and 3 and scores 4, 5, and 6 on the 6-point rubric and scores 1 and 2 and scores 3 and 4 on the 4-point rubric.

Using either scoring rubric, the first consideration for a reader is "How well has the student performed the task?" or, for example, "How *descriptive* is this paper? To what extent has the student done the job he or she has been asked to do?" (Alabama State Department of Education, undated d). This alone does not constitute the student's score (since this would be a primary trait score). Instead, the reader must evaluate the paper as a whole for Audience and Organization and Clarity (as defined by the rubric) as well. Although the ultimate score is weighted heavily toward mode considerations,

the "on mode" aspects of the paper are not the only aspects considered during scoring. A paper that only minimally achieves the task could conceivably be evaluated as a mid-range paper due to extremely strong elements within the Organization/Clarity section of the rubric. When all of these considerations have been weighted, readers must decide—"Is this paper a high paper, a mid paper, or a low paper?" Then, evaluating all elements (as described above), a focused holistic score from "1" to "4" is assigned.

Since the overriding philosophy of the Alabama Direct Assessment of Writing is to improve writing instruction (Alabama State Department of Education, undated d), materials and assistance supporting instruction are available. Instructional packets have been developed and provided to writing teachers for both Grades 5 and 7. Alabama State Department of Education professional staff provide assistance and inservice support when requested.

National Direct Writing Assessments

Since educators and the public began depending on large-scale standardized tests in the 1950s, direct assessment of student writing has been neglected. The late 1970s and early 1980s ushered in a resurgence in the assessment of student writing skills through the use of student-generated writing samples (Meredith & Williams, 1984). Supporting this trend was Odell, Cooper, and Courts (1979) in a National Council for Teachers of English publication that concluded that the direct method of assessing writing is the most appropriate. This trend was also supported by a Carnegie Foundation study (Boyer, 1983) that concluded that while writing is the most important skill taught in school, it is also the most neglected.

Direct assessments of writing were conducted at state and national levels during the 1980s. The National Assessment for Educational Progress (NAEP) writing prompts for a direct assessment of student writing were given to a national sample of students in 1984 and 1988 and again in 1990 (Gentile, 1992). South Carolina began the direct assessment of writing in 1981 (Meredith, 1984). Texas gave its first direct assessment of writing in 1983 (Sachse, 1984). Maryland began planning its direct assessment of writing program in 1978 which culminated with the assessment of 112,000 7th and 9th grade students (Hermann & Williams, 1984). A direct assessment of writing became part of Illinois' program in 1983 (Chapman, Fyans, & Kerins; 1984). Connecticut's direct assessment of writing began in the early 1980s and followed the NAEP's model (Baron, 1984). The Hawaii State Board of Education began using a direct assessment of writing in 1983, and by 1987 it had expanded to include grades 3, 6, 8, and 10 (Hawaii State Department of Education, 1989).

Quellmalz (1984) uses the research available up to the time her article was written to identify the elements of successful large-scale writing assessment programs. One of her four elements is "The assessment provides information that will be used at the classroom and system levels to improve writing skill" (p. 29). She also alludes to the use of writing assessments for influencing teachers' classroom practice.

The widespread implementation of direct assessments of writing as well as other forms of performance assessment during the 1980s raised concerns in some quarters about the validity of these efforts (Rothman, 1995). Fortunately, the widespread implementation of direct assessments of writing and performance assessment provided ample opportunity to study their validity as well as their utility. The next section explores some of the issues raised in some of these validity studies.

The Validity of Direct Assessments of Writing

It is well-known that tests that measure writing with multiple-choice items have questionable construct validity (e.g., Spivey, 1990). Having students pick a correct language usage from among options is not a good indicator of how well someone can write. The increased use of direct writing assessment methods in the 1980s provided researchers with many opportunities to explore the validity of these assessment methods. Studies are available that assess the validity of performance assessment methods in general and direct writing assessment in particular.

The face validity of direct writing assessment over multiple-choice tests is obvious. That is, measuring one's ability to write by having him or her write

appears to have more validity than having the student answer questions about writing. The real validity questions about direct writing assessments (and performance assessment in general) are whether or not scoring can be objective and reliable. The validity of a measure is constrained by its reliability. In a Michigan study of a locally developed writing assessment given to 642 sixth-grade students, Burger and Burger (1994) found that the interrater reliability was .69 and the writing assessment results were moderately related (.37 to .48) to Comprehensive Tests of Basic Skills, Fourth Edition, and Michigan Educational Assessment Program subscores. They conclude that "the moderate correlations representing criterion validity for the performance measures suggest the two types of assessment are measuring something in common, but they most certainly are not measuring identical skills" (p. 13). In another study of the validity of writing assessment, Farr and Jongsma (1993) found substantially higher interrater reliabilities (.84 to .97). They also concluded that their results "provide strong evidence for both convergent and discriminant validity" (p. 89). That is, the scores are highly related for the same or similar constructs that are being measured and are less related, as would be expected for different constructs.

A key element to the validity of any direct writing assessment is the reliability with which the writing samples are scored. In the two previous examples (i.e., Burger & Burger, 1994 & Farr & Jongsma, 1993), the interrater reliabilities were quite different. Gearhart, Herman, Novak, Wolf, and Abedi (1994) suggest that a key may be in the training of the raters. In an intense investigation into the validity of one writing prompt and two scoring rubrics (a holistic and holistic/analytic combination), Gearhart et al. examined interrater reliability over multiple grades (1 through 6) for a two-year period. They included five raters with varying backgrounds. Using generalizability theory, they concluded that the amount of variance due to raters was "almost negligible" (p. 21) but the generalizability coefficients ranged from .47 to .68. They concluded that "adequate levels of reliability for most scales of either rubric could be attained by doubly scoring each essay and aggregating the results" (p. 34). Abedi, Gearhart, Herman, Novak, and Wolf (1994) report similar findings.

Another validity dimension often criticized for performance-based assessments is their potential inequitable effects on different populations of students (Darling-Hammond, 1994). Welch and Miller (1995) found evidence that direct writing assessment items did function differentially for different populations of students. They conclude that the "large differences in effects between the male/female and the African American/White analyses, despite greater mean differences in the latter and even after adjusting for sample size, also suggest group-related construct differences between item types" (p. 175). Darling-Hammond argues that "for all the promise of more authentic and performance-based forms of assessment, their value depends as much on how they are used and what support for learning accompany them as on the new technologies they employ" (p. 25).

These validity studies suggest that direct assessments of writing and performance assessments in general have more than adequate validity for instructional uses. However, they may not be adequate for use in high stakes situations such as student graduation or promotion examinations or for school or system accreditation decisions. Additional research would need to be completed before they are used for such purposes.

The Impact of Assessments on Curriculum

It has long been assumed that curricular changes could be influenced by large-scale assessments. There is evidence that the earliest example of system-wide assessment in America was prompted by this belief. In 1845, Horace Mann, secretary to the Massachusetts State Board of Education, asked the Boston School Committee to convert from oral examinations to written examinations (Rothman, 1995). The thinking behind this request was to increase the pressure for change by demonstrating that students were deficient in the knowledge being measured by the examination. While this testing program accomplished exactly what Mann had hoped, his legacy is the large-scale assessments of today that are still influencing curriculum. Studies are available supporting the impact of testing on the curriculum in general and of the influence direct writing assessment has on the teaching of writing.

The concept of testing influencing the curriculum has added the phrase "measurement-driven reform" to our vocabulary (Noble & Smith, 1994). Popham (1987) described how assessment results directed teachers' attention to the content of the assessments. Cabay, Noble, and Smith (1994) used multiple case studies to demonstrate how the curriculum changes when testing mandates change. Some assessment programs such as the Arizona Student Assessment Program (ASAP) and that spawned by Florida's Educational Accountability Act of 1971 were designed from the outset to influence instruction. Some feel that the more high stakes the test, the more it influences instruction (e.g., Linn, 1993). In fact, Linn warns that mandated testing programs sometimes have unintended outcomes that may not be entirely straightforward. Resnick and Resnick (1989) were more direct in asserting that: (a) you get what you assess, (b) you do not get what you do not assess, and (c) you should build assessments directed at what you want educators to teach. The third point was logically concluded from the first two which are based on research findings (Noble & Smith, 1994).

The general conclusions drawn about the impact of testing on instruction are not as well documented for performance assessments and direct assessments of writing in particular. However, in a recent *NAEPFACTS* newsletter, it was not only reported that direct assessment of student writing encouraged more writing instruction, but students whose teachers emphasized certain process writing skills exhibited better-developed writing skills on the NAEP writing assessment (National Assessment of Educational Progress, 1996). This lack of evidence of the impact of direct writing assessments on writing instruction suggests additional research should be completed on this question.

Summary of Literature

Alabama's direct assessment of writing is part of a national trend that has increasingly used performance assessment as part of the student assessment program. Assessing students' ability to write by examining examples of student writing has been found to have strong face and content validity and adequate reliability for instructional purposes. Some questions still remain about the validity of direct writing assessment for promotion, graduation, or accountability decisions. The literature provides much support for the idea that we teach what we test. However, little direct evidence is available to extend this to the assessment of student writing samples. While this result can be established by logically extending the findings from other areas, data needs to be examined to establish this directly.

Psychometric Analysis

Psychometric analyses of the Alabama Writing Assessment were conducted to address several key questions. These were

1) What are the patterns of performance for students on the Alabama Writing Assessment from 1992 through 1995? To what extent do these patterns vary as a function of the type of writing, the year of assessment, or the grade level of the student?

2) To what extent is there evidence of consistency between multiple raters of the writing samples?

3) To what extent is there evidence of stability of performance between 5th and 7th grade for students required to do the same type of writing each year? To what extent does the prompt mediate such an effect?

Sample

To address the research questions for this study a number of different data sets were constructed from the core data sets for the 1992, 1993, 1994, and 1995 assessments. Variations in the constitutions of the data sets precluded common aggregations. Four differences are most notable.

First, the 1995 data included the summated score across two raters but did not include scores for each independent rater. This prohibited analyses regarding the consistency of raters for 1995. Also, in 1995 a new four-level scoring rubric replaced the six-level scoring rubric used in 1992 through 1994. This necessitated dichotomizing outcomes (at or above grade-level, below grade-level) in order to be able to compare all data sets. Third, no 7th grade data were available for 1992. Finally, four types of writing are assessed in 7th grade, whereas only three types of writing were assessed in 5th grade. Therefore, analyses comparing 5th and 7th grade performance are limited to students who in 7th grade took one of the three prompts that were also used in 5th grade (narrative, descriptive, expository).

Data Sets

Annual Data Sets. The primary data sets were the annual test results from grades 5 and 7 for each year the particular assessment was given (1992-1995 for 5th grade, 1993-1995 for 7th grade). Primary data sets were used to generate descriptive data (means, frequencies) and for comprising subsets needed for the inferential analyses (research questions).

Cross-sectional Subsets. Three independent samples of 1000 subjects from each of the primary annual data sets were randomly selected and merged. This yielded three subsets of 7000 subjects each. These data sets were used to address research questions 1 and 2.

Longitudinal Subsets. In order to address longitudinal research question (#3), four independent subsamples were constructed of 1000 students for whom both 5th and 7th grade assessments were available. Two subsets were from the 1992/1994 cohort and two from the 1993/1995 cohort.

Results

Research Question 1. Due to differences in scoring rubrics between the 1992-1994 and 1995 assessments (changing from a 6 point to a 4 point rubric), it was necessary to dichotomize the summated holistic ratings (acceptable at or above grade-level performance vs. unacceptable below grade-level performance) in order to provide a common metric for each year. Table 1 presents the average dichotomized ratings for each year, grade-level, and prompt. Scores in this table represent the proportion of students receiving a rating of "acceptable" on the assessment.

Table 1

Proportion of Students Scoring at Acceptable Levels or Above on Writing Performance

Grade	Year	Prompt				
		Overall	Narrative	Descriptive	Expository	Persuasive
5	92	.229 (n=50546)	.450 (n=17650)	.165 (n=17226)	.042 (n=15670)	NA
	93	.197 (n=52103)	.391 (n=17517)	.084 (n=17811)	.117 (n=16775)	NA
	94	.168 (n=50593)	.360 (n=17024)	.102 (n=17130)	.041 (n=16539)	NA
	95	.200 (n=49596)	.228 (n=17372)	.200 (n=16245)	.169 (n=15979)	NA
7	93	.135 (n=51295)	.286 (n=13817)	.130 (n=11334)	.039 (n=13322)	.075(n=12822)
	94	.108 (n=51837)	.195 (n=13114)	.090(n=12664)	.070 (n=13200)	.076 (n=12859)
	95	.265 (n=52102)	.377 (n=13225)	.230 (n=12764)	.233 (n=13300)	.218 (n=12813)

In both 5th and 7th grade, narrative writing resulted in the greatest proportion of students performing at acceptable levels. The expository writing assessment was most frequently the most difficult type of writing prompt at both grade levels. These trends were confirmed using inferential statistics (ANOVA) with the cross-sectional data subsets. Specifically, statistically significant differences were found between average proportions for all pairs of prompts (types of writing) for 5th grade and between the narrative prompt and all other prompts in 7th grade.

Results in Table 1 indicate that between 1992 and 1995 fewer than 40% of Alabama's 5th and 7th grade students displayed writing skills at acceptable levels. For descriptive, expository and persuasive writing, fewer than 25% of students yielded writing samples of acceptable quality.

Table 1 also reveals notably higher performance for 7th graders in 1995 than in previous years. The magnitude of the discrepancies may be a function of true educational improvement resulting from greater attention to writing or statistical artifacts associated with moving from a six-point to a four-point scoring rubric. As no parallel discrepancy was found for 5th grade, it is unlikely that the improvement was due to a change in the scoring rubric. As the Psychological Corporation was responsible for piloting and scoring the writing prompts every year in 7th grade, it is unlikely that all writing prompts for 1995 were systematically easier or grading was systematically more liberal in 1995 than in previous years. Examination of the 1996 data, when available, should provide evidence as to the stability of the improvement.

Additional Finding. Table 2 provides additional descriptive information concerning the performance of Alabama students on the Alabama Direct Assessment of Writing for 1995. These data were derived from the 1995 Report of the Chief State Education Officer. Table 2 displays the unweighted averaged percentages of students within each school system performing at or above acceptable levels on the 5th and 7th grade assessments for eight demographic categories (two levels of size and four levels of socioeconomics) in Alabama schools. As can be seen, the general pattern is that the proportion of success declines as socioeconomic status declines. However, large below average socioeconomic grade seven schools yielded the highest success rate of all large school categories, and the second highest success rate of any category (second only to the wealthiest small schools). Since these results are based on the total sample, we can assume this is a real finding and not an artifact.

BEST COPY AVAILABLE

Table 2

Percentages of Students Within Clusters
Above the Acceptable Level on the
1995 Alabama Direct Writing Assessment

	Grade	
Cluster	5th	7th
Small, High Socioeconomic	40.94	55.37
Small, Above Average Socioeconomic	21.84	33.48
Small, Below Average Socioeconomic	20.32	25.94
Small, Low Socioeconomic	11.41	14.88
Large, High Socioeconomic	23.19	33.06
Large, Above Average Socioeconomic	20.38	26.40
Large, Below Average Socioeconomic	16.18	42.65
Large, Low Socioeconomic	12.61	16.21

Research Question 2. As part of the assessment process, a minimum of two raters scored every writing sample. When ratings between two raters differed by more than one point on the holistic rating scale, a third rater scored the writing sample. Inter-rater reliability data was available for 5th grade data set for the years 1992 through 1994. Discrepancies between the two primary raters ranged from 0 to 4 points. Table 3 reports the average discrepancy between the two initial raters, the correlation (r) between the raters, and the percentage of variance (r^2) in scores that was shared by the two raters.

Therefore, rater pairs disagreed on the average by less than one-half point in each year. In addition, the most common outcome between rater pairs for each year was exact agreement. During each year, more than 50% of writing samples were rated identically by the two primary raters.

To confirm the stability of scores between raters a generalizability theory approach was used with 5th grade data from selected cross-sectional data subsets. No statistically significant differences between raters were found in any of these random cross-sectional samples. When the consistency of raters was examined by prompt and by year of the test, only one statistically significant interaction was found in the three samples, a prompt-by-rater effect. This suggests that, to some degree the consistency of ratings may vary as a function of the type of writing that is being graded. However, it should be noted that as the sample sizes of these analyses were very large (n >2800), the statistical procedures will detect very small differences as statistically significant. Inter-rater reliability of the 5th grade scores for 1992-1994 was acceptable. Data were not available to make conclusions for 7th grade or for 1995.

Table 3

Summary of Discrepancy, Correlation, and Shared Variance by Year

Year	Average Discrepancy	Correlation (r)	Shared Variance (r^2)
1992	.475	.685	.469
1993	.423	.637	.406
1994	.415	.654	.428

Research Question 3. To examine changes in individual writing performance between grades 5 and 7, students who had been required to do the same type of writing in both grades were identified from the longitudinal data sets. Matching prompts were needed in order to remove the confounding associated with differences in the difficulty between writing prompts for the assessment of change. Furthermore, since the 1995 scoring rubric was different from the 1993 rubric, the ratings from these years were dichotomized in order to allow comparison. As a common rubric was used in 1992 and 1994, there was no need to dichotomize the scores. Furthermore, as four types of writing (prompts) were assessed in 7th grade and three types in 5th grade, analyses are limited to those students who did narrative, descriptive, or expository writing in 7th grade.

Statistically significant differences due to time were found as a result of repeated measures analyses of variance. However, the nature of the significant difference varied with the cohort. The 1992/94 cohort demonstrated better performance overall in 5th grade than in 7th grade. The 1993/95 cohort demonstrated better performance overall in 7th grade. Further analyses showed that the most notable difference between 5th and 7th grade for the 1993/95 cohort occurred as a result of substantial gains in descriptive writing between 1993 and 1995. Significant interactions between prompt and time were noted for the 1992/94 cohort. This was due to expository writing performance being the only type of writing that was better in 7th grade then in 5th for this cohort. No prompt by time interaction was found for the 1993/95 cohort.

Empirical Evaluation

The empirical evaluation included both qualitative and quantitative methods based on interviews and surveys of 5th and 7th grade writing teachers. The interviews were conducted with teachers from purposefully selected schools. The surveys were sent to teachers from randomly selected schools. The evaluation is described in terms of its qualitative and quantitative aspects.

Qualitative Study

The first phase of the empirical evaluation was interviews with writing teachers. This began with the determination of three groupings of schools—those whose performance on the Alabama Direct Assessments of Writing has been consistently high, improving, and consistently low. A school was assigned to the consistently high group when at least 70% of its students scored in the "Good" or "Elaborated" categories (Levels III and IV on the 4-point scale) on each assessment. It should be noted that these categories corresponded to Levels IV, V, and VI on the 6-point rubric used between the years 1992 and 1994. A school was assigned to the consistently low group when less than 30% of its students were above Level III on each assessment. A school was assigned to the improving group when a school demonstrated improvement in the percentage of students above Level III on each occasion the assessment was given and the overall improvement exceeded 30 percentage points. The schools were selected in consultation with the Alabama State Department Language Arts Specialist, Ms. Ann Moody, who worked with the Alabama Direct Assessment of Writing Program from its inception. Five schools were selected from each group and permission was obtained from the superintendent for that school to participate.

Two experienced doctoral-level educators conducted the interviews. They contacted the principal, set a date, and conducted an onsite visit. They asked to interview all teachers who taught either 5th or 7th grade writing in each school. Some teachers were not available, and participation in the interview was voluntary. The interviews, approximately 15-30 minutes in length, were based on an Interview Protocol (see Appendix A) developed for the evaluation. The items were based on a review of the literature and input from Alabama State Department of Education personnel and a small sample of writing teachers. The instrument was piloted and reviewed by Alabama State Department personnel.

Qualitative Results. The interview data were analyzed both quantitatively and qualitatively. The responses were tabulated and are presented by group as frequencies and percentages. The data were analyzed using descriptive procedures due to the purposeful sampling of the schools. This means that inferential statistics were not appropriate. Table 4 provides a summary of the demographic data by group (low, improving, and high).

The qualitative results also included the impressions of the interviewers based on their interviews with the 5th and 7th grade writing teachers. The interviewers were not aware of the bases on which the schools were selected to participate in the interviews or that there were three identified groups. The interviewers submitted their impressions to the principal investigators identified only by school. The principal investigators reinterpreted their impressions in terms of the three groups.

One of the first characteristics that seemed to divide the groups was their responses to the consent form that each teacher was asked to sign. Teachers from two of the upper schools did not bother to even read the form, but listened to the interviewer's brief explanation and signed it. The other extreme was at a lower-achieving school where the teachers read, and in some cases reread, the form and seemed very concerned about signing it. Most of the remaining teachers read the form before signing it. Another observation by one of the interviewers was the high level of personal articulation and confidence of teachers from two of the upper achieving schools. They appeared to be very confident and articulate teachers, especially with respect to the teaching of writing.

Table 4

Description of Interviewees

Characteristic		Group: Low number (%)	Group: Improving number (%)	Group: High number (%)
Level of Certification:	Bachelors	7 (41%)	8 (47%)	3 (20%)
	Masters	10 (59%)	8 (47%)	10 (67%)
	Specialist	0 (0%)	1 (6%)	2 (13%)
	Total	17	17	15
Grade Level Taught:	Grade 5	6 (35%)	7 (41%)	7 (47%)
	Grade 7	11 (65%)	10 (59%)	8 (53%)
	Total	17	17	15
Years Taught Writing:	One year or less	1 (6%)	2 (12%)	2 (13%)
	Two to five years	5 (29%)	4 (24%)	4 (27%)
	Six to ten years	3 (18%)	3 (18%)	4 (27%)
	More than ten years	8 (47%)	8 (47%)	5 (33%)
	Total	17	17	15

It seemed that a majority of the upper achieving schools had a teacher who served as a leader in developing a school plan for teaching writing. These teachers indicated that they place great value on writing in general and have developed an organized systematic approach that is year-long with some review just prior to the assessment. They prepared students in a variety of ways, including writing from the beginning of the year in a systematic effort, sharing, and, in some cases, expanding the scoring criteria (rubrics), reviewing samples from the state, having the students score some examples using the rubrics, and reviewing just prior to the assessment.

The majority of schools in the lower-achieving group and one school in the improving group were not as consistent in the development of an overall plan to help students become better writers. Some teachers in these schools were unsystematic about their approaches to teaching students to write. Some of these teachers expressed knowledge of the rubric, but were not able to provide a clear picture when asked by the interviewer to explain further.

Teachers from all groups seemed to share the opinion that the teaching of writing would suffer if the Alabama Direct Assessment of Writing was eliminated. Many claimed that they would continue, but "other teachers" needed the pressure of the writing assessment to get them to find the large amounts of time necessary to teach writing. A small minority claimed that they could teach writing better without the constraints of the writing assessment. Table 5 presents a summary of the interview questions by item.

In addition to the yes-no responses presented in Table 5, the teachers who answered "yes" were asked to describe their "yes" answers more fully. The most revealing of these questions was "Have you used the Alabama scoring rubric in your teaching of writing?" If a teacher answered "yes," he or she was asked to describe how it was used. The recorded responses to this follow-up question were examined using content analysis and computing frequencies and percentages of responses by content category for each group. Table 6 shows the categories and responses by group. As can be seen, the majority of faculty in the "Improving" and "High" groups used rubrics for instructional planning purposes while teachers in the "Low" performing group were more diffuse in their applications of the prompts.

BEST COPY AVAILABLE

Table 5

Number and (Percentages) Responding "Yes" to Interview Items

Item	Group		
	Low number (%)	Improving number (%)	High number (%)
Have you been directly involved in the development of a state direct writing assessment (e.g., advisory committee, content and bias review)?	3 (18%)	1 (6%)	1 (7%)
Have you administered a direct writing assessment in your school?	17 (100%)	17 (100%)	15 (100%)
Are you familiar with the state's direct writing assessment scoring rubric?	14 (82%)	17 (100%)	15 (100%)
Have you used the Alabama scoring rubric in your teaching of writing (e.g., for planning purposes, for scoring of student writing)?	14 (82%)	15 (88%)	15 (100%)
Do you do anything special (other than you would do ordinarily) to prepare students for the Alabama Direct Writing Assessment?	13 (76%)	12 (71%)	11 (73%)
If the Alabama Direct Writing Assessment were no longer required, do you believe writing instruction would change?	16 (94%)	9 (60%)	13 (93%)
Do you have any recommendations for improvingthe Alabama Direct Writing Assessment?	15 (88%)	14 (100%)	13 (87%)

Quantitative results. The second phase of the empirical evaluation was a survey of 5th and 7th grade writing teachers in a state-wide random sample of schools. Thirty schools that included a 5th grade and 30 schools that included a 7th grade were randomly selected. Once a school system was selected, it was removed from consideration. Thus, a total of 60 school systems were selected. The schools were not assigned to groups as in the interview phase of the evaluation; however, the mean performance on the Spring, 1995, Alabama Direct Assessment of Writing was computed for each school selected. Permission was sought from the appropriate system superintendent to survey the teachers in that school. Permission was granted in 49 of the 60 cases requested. Six local superintendents sent back "no" responses to the requests and five did not respond at all. Thus, 49 gave permission for a school in their district to participate.

Packets were sent to the principals of the 49 schools with a letter explaining the purpose of the survey. The packets included a copy of the signed permission form from the local superintendent and five teacher packets, each in a sealed envelope. The principal was asked to give a teacher packet to each teacher who taught either 5th or 7th grade writing. If there were more than five teachers who taught writing, then the principal was to give the five packets to five randomly selected teachers. Each

Table 6

How the Rubric Was Used by Interviewees

Use Category	Group		
	Low number (%)	Improving number (%)	High number(%)
Planning	6 (25%)	11 (85%)	5 (62%)
Teachers use in grading papers	3 (12%)	1 (8%)	0 (0%)
Students use in grading own papers	3 (12%)	0 (0%)	0 (0%)
Teach students actual rubrics	5 (21%)	0 (0%)	0 (0%)
Explain rubrics generally but do not teach directly	2 (8%)	1 (8%)	1 (25%)
Students use in peer editing	3 (12%)	0 (0%)	0 (0%)
Use papers graded using rubrics as part of class grade	2 (8%)	0 (0%)	0 (0%)
Totals	24	13	8

teacher packet included a letter from one of the principal investigators that gave instructions and explained the purpose of the survey, a consent form, a survey form, and a pre-addressed and stamped envelope for return directly to UAB.

The surveys (see Appendix B) included some items from the Interview Protocols and other items to gain a broader perspective of the perceptions about the Alabama Direct Assessment of Writing Program. As with the Interview Protocols, the items were based on the review of the literature, input from Alabama State Department of Education personnel, and a small sample of writing teachers. The instrument was pilot tested and reviewed by Alabama State Department personnel.

A total of 50 surveys were returned. Only teachers who responded that they did teach either 5th or 7th grade writing were included in the analyses. The returns from teachers represented 30 schools of the original 49 schools sent survey forms. No follow-ups were conducted due to the lateness of the school year. Data were analyzed similarly to the Interview data with both quantitative and qualitative analyses being conducted. However, only one item (Item 8) requested an open ended response. Table 7 includes a summary of the demographic results and items about their involvement in the teaching of writing.

As with the equivalent interview question, the item, "Have you used the scoring rubric in your teaching of writing?" requested the teachers to describe how the rubric was used when they answered "yes." A content analysis was performed on these responses using the same content categories that were used in the analysis of the interview item reported in Table 6. The results of this analysis are reported in Table 8.

Items 9 through 20 were based on a 5-point "Strongly Disagree" to "Strongly Agree" Likert scale. For each of these items, the frequency and percentage response distributions are provided as well as the mean and standard deviation for each item. For convenience of interpretation, the items are listed in order of level of agreement based on the computed means. The items with the largest means or strongest agreement are listed first. These results are presented in Table 9.

BEST COPY AVAILABLE

Table 7

Demographic and Involvement Variables from Survey

Item		Number and (Percentage)
Level of Alabama Teaching Certification:	Bachelors	22 (44%)
	Masters	23 (46%)
	Specialist	3 (6%)
	No answer	2 (4%)
	Total	50
Grade writing is taught:	Grade 5	27 (54%)
	Grade 7	23 (46%)
	Total	50
Number of Years Experience Teaching Writing:	One year or less	2 (4%)
	2-5 years	23 (46%)
	6-15 years	12 (24%)
	More than 15 years	12 (24%)
	Total	50
Have you been directly involved with the development of a statedirect writing assessment?	Yes	1 (2%)
	No	49 (98%)
Have you administered a direct writing assessment in your school?	Yes	41 (82%)
	No	9 (18%)
Are you familiar with the state's writing assessment scoring rubric?	Yes	35 (70%)
	No	13 (26%)
	Missing	2 (4%)
Have you used the scoring rubric in your teaching of writing?	Yes	20 (40%)
	No	29 (58%)
	Missing	1 (2%)

Table 8

How the Rubric Was Used by Survey Respondents

Category	Number	Percent
Teacher Grading Papers	9	32.1
Planning	8	28.6
Explained to Students	4	14.3
Taught to Students	3	10.7
Use Assessment as Part of Class Grade	2	7.1
Students Grading Own Papers	1	3.6
Students Use in Grading Peers' Papers	1	3.6
Total	28	100.0

Note. These results represent responses from 22 teachers. Some teachers had more than one response.

Table 9

Responses to Survey Items

Item		Frequency and Percentage Distributions						Mean	SD
		Strongly Disagree				Strongly Agree			
		1	2	3	4	5	NA/DK		
I have made specific efforts to align my writing instruction with the state's writing assessments.	n:	3	1	7	10	26	3	4.17	1.17
	%:	6.4	2.1	14.9	21.3	55.3			
I have received support and encouragement from my school administrators related to the state mandated writing assessments.	n:	6	3	6	9	22	4	3.83	1.44
	%:	13.0	6.5	13.0	19.6	47.8			
As a result of the statewide direct writing assessments, my teaching of writing has improved.	n:	1	5	9	18	13	4	3.80	1.05
	%:	2.2	10.9	19.6	39.1	28.3			
I have improved instruction and assessment in my English/language arts classes as a result of my state's implementation of the direct writing assessments.	n:	3	8	9	13	12	5	3.51	1.25
	%:	6.7	17.8	20.0	28.9	26.7			
The state direct writing assessment is based on sound principles of writing.	n:	8	1	13	10	15	3	3.49	1.41
	%:	17.0	2.1	27.7	21.3	31.9			
Preparing my students to take mandated writing assessments has had a positive effect on student writing skills.	n:	8	2	12	13	12	3	3.4	1.38
	%:	17.0	4.3	25.5	27.7	25.5			
I believe the state mandatory writing assessments have encouraged teachers of writing toward a common curriculum.	n:	4	7	10	17	8	4	3.39	1.20
	%:	8.7	15.2	21.7	37.0	17.4			
I like my state's efforts to improve instruction through their implementation of mandatory assessments in writing.	n:	7	2	19	8	11	3	3.30	1.30
	%:	14.9	4.3	40.4	17.0	23.4			
As a result of the statewide direct writing assessments, I believe that the curriculum taught in English/language arts is becoming more uniform statewide.	n:	5	5	8	17	6	9	3.34	1.24
	%:	12.2	12.2	19.5	41.5	14.6			
Teaching my students to use the scoring rubrics for the state mandated writing assessments has increased student confidence in their writing.	n:	2	5	10	5	1	27	2.91	1.00
	%:	8.7	21.7	43.5	21.7	4.3			
I believe that the state mandatory writing assessments have had a negative impact by narrowing the variety of writing taught and practiced in the classroom.	n:	13	12	11	4	8	2	2.62	1.41
	%:	27.1	25.0	22.9	8.3	16.7			
I believe that the state mandated assessment results are an accurate reflection of student writing performance.	n:	17	7	9	8	4	5	2.44	1.39
	%:	37.8	15.6	20.0	17.8	8.9			

BEST COPY AVAILABLE

As previously noted, the mean performance of students in each school on the 1995 Alabama Direct Assessment of Writing was computed and added to the data base. This was used in a regression analysis to determine its relationship to the responses to the items noted in Table 9. Specifically, the mean school performance on the Alabama Writing Assessment was regressed on the 12 Likert-scaled items shown in Table 9. The overall Writing Assessment variance accounted for by the teachers' responses to the 12 items was .43. To further understand the contribution of each item to this relationship, the structure coefficients or correlation of each item with the Writing Assessment performance predicted by the regression equation was computed. These results are shown in Table 10.

Table 10

Correlations of Predicted Writing Performance with Item Responses

Item	Correlation with Predicted Achievement
Teaching my students to use the scoring rubrics for the state mandated writing assessments has increased student confidence in their writing.	.77
I like my state's efforts to improve instruction through their implementation of mandatory assessments in writing.	.65
I have made specific efforts to align my writing instruction with the state's writing assessments.	.64
As a result of the statewide direct writing assessments, my teaching of writing has improved.	.64
I believe that the state mandated assessment results are an accurate reflection of student writing performance.	.57
I have received support and encouragement from my school administrators related to the state mandated writing assessments.	.55
I have improved instruction and assessment in my English/language arts classes as a result of my state's implementation of the direct writing assessments.	.49
The state direct writing assessment is based on sound principles of writing.	.39
I believe that the state mandatory writing assessments have had a negative impact by narrowing the variety of writing taught and practiced in the classroom.	.37
I believe the state mandatory writing assessments have encouraged teachers of writing toward a common curriculum.	.17
Preparing my students to take mandated writing assessments has had a positive effect on student writing skills	-.16*
As a result of the statewide direct writing assessments, I believe that the curriculum taught in English/language arts is becoming more uniform statewide.	-.44

* The relationship for this item is based on insufficient data and is not stable.

All but two of the relationships denoted in Table 10 are statistically significant at .01 level. The one marked with an asterisk is not significant, but due to the large number of teachers who omitted this item, it is not stable anyway. The other one that is not significant is the .18 (p=.16). The higher the correlation reported in Table 10, the more that particular statement by teachers is related to student achievement on the Alabama Direct Assessment of Writing. Most of the relationships were positive, meaning that the more strongly teachers agreed with the statement, the better students performed on the writing assessment. However, there was a significant negative relationship between the last statement and student performance. That is, when teachers tended to believe that the English/language arts curriculum was becoming more uniform as a result of the statewide direct writing assessments, their students tended to perform more poorly on the writing assessment.

Summary and Conclusions

The evaluation of the Alabama Direct Assessment of Writing Program illuminated a number of key issues and outcomes. These findings extend the existing knowledge base about state-level performance assessments as well as provide useful information to decision-makers in Alabama. The following key conclusions were drawn from the evaluation:

- The majority of 5th and 7th grade students in Alabama do not perform at acceptable standards in any of the major types of writing.
- Considerable gains in writing performance were evident in 7th grade between 1994 and 1995 across all types of writing. This improvement appears to reflect improved educational opportunities for students rather than statistical or measurement artifacts.
- Moderate but sufficient agreement was evident between the two principle scorers of each writing response. While the interrater reliabilities would not be adequate to make high-stakes decisions about individual students, they are sufficiently strong to aggregate judgments. The state's scoring system uses Gearhart et al.'s (1994) recommendation for ensuring adequate reliability in that the scores for the independent raters are combined when determining performance.
- Teachers are in considerable agreement that the state-mandated writing assessment has positively influenced the teaching of writing and that the emphasis on student writing would diminish if the writing assessment were eliminated.
- High-performing schools tended to place great value on writing and applied year-long systematic approaches to writing instruction. The use of scoring rubrics for instructional planning was among the strategies that appear to differentiate the high-performing and improving schools from the low-performing schools.
- In the lower-achieving schools, the teachers did not appear to have the level of skill necessary to develop a cohesive writing curriculum effectively using the writing assessment results. This suggests that they will likely need assistance to improve student writing skills.
- The more positive that teachers' perceptions of the State Direct Assessment of Writing are, the higher the average student performance on the writing assessment at the school tends to be.

The basic findings of the evaluation support the psychometric integrity of the writing assessment and its role in encouraging the teaching of writing in Alabama schools. The evaluation also found that for the Alabama Direct Assessment of Writing to be of greatest benefit, many teachers in the schools that have been less successful on the writing assessment must be provided with additional support to learn how to deal with the information they gain more effectively. The writing assessment has a positive overall influence on the ability of students in Alabama to communicate effectively through the written word.

References

Abedi, J., Gearhart, M., Herman, J. L., Novak, J. R., Wolf, S. A. (1994). *Toward the Instructional Utility of Large-Scale Writing Assessment: Validation of a New Narrative Rubric.* (CSE Tech. Rep. No. 389). Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Alabama State Department of Education (undated a). *Alabama Direct Assessment of Writing: Grade Seven, Test Administrator's Manual.* Montgomery, AL: author.

Alabama State Department of Education (undated b). *Alabama Direct Assessment of Writing: Grade Five, Instructional Packet.* Montgomery, AL: author.

Alabama State Department of Education (undated c). *Alabama Direct Assessment of Writing: Grade Seven, Instructional Packet.* Montgomery, AL: author.

Alabama State Department of Education (undated d). *Alabama Direct Assessment of Writing: Philosophy and Information.* Montgomery, AL: author.

Baron, J. B. (1984). Writing assessment in Connecticut: a holistic eye toward identification and an analytic eye toward instruction. *Educational Measurement: Issues and Practice, 3*(1), 27-28.

Boyer, E. L. (1983). *High school, a report on secondary education in America* (Carnegie Foundation for the Advancement of Technology). New York: Harper and Row.

Burger, D. L., & Burger, S. E. (1994). Determining the validity of performance-based assessment. *Educational Measurement: Issues and Practice,* 9-15.

Cabay, M., Noble, A. J., Smith, M. L. (1994). *What Happens When the Test Mandate Changes? Results of a Multiple Case Study.* (CSE Tech. Rep. No. 380). Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chapman, C. W., Fyans, L. J. Jr., Kerins, C. T. (1984). Writing assessment in Illinois. *Educational Measurement: Issues and Practice, 3*(1), 24-26.

CTB Macmillan/McGraw-Hill (1992). *Alabama Direct Assessment of Writing: Grade Five, Test Administrator's Manual.* Monterey, CA: Author.

Darling-Hammond, L. (1994). Performance-Based Assessment and Educational Equity. *Harvard Educational Review,* 64(1), 5-30.

Farr, R., & Jongsma, E. (1993). The convergent/discriminant validity of integrated reading/writing assessment. *Journal of Research and Development in Education,* 26(2), 83-91.

Gearhart, M., Herman, J., Novak, J., Wolf, S., and Abedi, J. (1994). *Toward the Instructional Utility of Large-Scale Writing Assessment: Validation of a New Narrative Rubric.* (CSE Tech. Rep. No. 389). Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Gentile, C. (1992). *Exploring New Methods for Collecting Students' School-based Writing.* Washington, DC: National Center for Education Statistics.

Hammond-Darling, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review,* 64(1), 5-30.

Hawaii State Department of Education. (1989). *Program Analysis of Stanford Writing Assessment Grades 3, 6, 8, and 10. Spring 1987 and Spring 1988 Administrations.* (Report No. CS-212-459). Honolulu, Hawaii: Office of Instructional Services. (ERIC Document Reproduction Service No. ED 322 511)

Hermann, F., & Williams, P. L. (1984). Writing assessment in Maryland. *Educational Measurement: Issues and Practice, 3*(1), 23-24.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15*(1), 1-16.

Meredith, V. H. (1984). Writing assessment in South Carolina: past and present. *Educational Measurement: Issues and Practice, 3*(1), 19-21.

Meredith, V. H., & Williams, P. L. (1984). Issues in direct writing assessment: problem identification and control. *Educational Measurement: Issues and Practice*, 11-15.

National Assessment of Educational Progress. (1996). *NAEPFACTS*. Author, 1(3).

Noble, A. J., & Smith, M. L. (1994). *Measurement-Driven Reform: Research on Policy, Practice, Repercussion*. (CSE Tech. Rep. No. 381). Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Odel, L., Cooper, C., & Courts, C. (1979). Discourse theory: Implications for research in composing. In C. R. Cooper & L. Odell (Eds.), *Research on comparing: Points of departure*. Urbana, IL: National Council for Teachers of English.

Parker, R. I., Tindal, G., & Hasbrouck, J. (1991). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children*, 61-73.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68*(9), 679-682.

Quellmalz, E. S. (1984). Toward successful large-scale writing assessment: where are we now? where do we go from here? *Educational Measurement: Issues and Practice, 3*(1), 29-32.

Resnick, L. B., & Resnick, D. P. (1989). *Assessing the thinking curriculum: New tools for educational reform*. National Commission on Testing and Public Policy.

Rothman, R. (1995) Measuring Up : standards, assessment, and school reform, Jossey-Bass Publishers, San Francisco.

Sachse, P. P. (1984). Writing assessment in Texas: practices and problems. *Educational Measurement: Issues and Practice, 3*(1), 21-23.

Spivey, N. N. (1990). Transforming tests: Constructive processes in reading and writing. *Written Communication, 7*, 256-287.

Welch, C. J. & Miller, T. R. (1995). Assessing differential item functioning in direct writing assessments: problems and example. *Journal of Educational Measurement*, 32 (2), 163-178.

Williams, P. L. (1984). The application of direct writing assessments in five states. *Educational Measurement: Issues and Practice, 3*(1), 19.

Appendices

Appendix A
Interview Protocol

Alabama Direct Writing Assessment
Conducted by the Center for Educational Accountability
The University of Alabama at Birmingham
May 1996

For official use only: School Code:_____

Introduction: We are conducting an evaluation of the Alabama Direct Writing Program for the Alabama State Department of Education. The evaluation is sponsored by the SouthEastern Regional Vision for Education (SERVE), a federally funded regional educational laboratory. None of your responses will be linked with you individually. They will be used in combination with those of numerous other teachers in this state to make judgments about the Alabama Direct Writing Program. The evaluation has the full backing and support of the State Superintendent of Education and the Alabama State Department of Education.

1. Do you teach writing skills as part of your teaching duties? YES NO

   (If the answer is NO, terminate the interview as soon as feasible.)

2. At what level is your Alabama teacher certification? B A AA

3. Do you teach students writing at grade 5 YES NO
   or grade 7? YES NO

   (If the answer is NO for both grades 5 & 7, terminate interview.)

4. How long have you taught writing?
   YEARS_____

5. Have you been directly involved in the development of a state direct writing assessment (e.g., advisory committee, content and bias review)? YES NO

   If YES, in what capacity?_____

6. Have you administered a direct writing assessment in your school? YES NO

7. Are you familiar with the state's direct writing assessment scoring rubric? YES NO

   (You may share with them a copy. If their answer is NO, skip to Item 9.)

8. Have you used the Alabama scoring rubric in your teaching of writing (e.g., for planning purposes, for scoring of student writing)? YES NO

If YES, please describe how you have used the rubric.

_____

_____

_____

_____

9. Do you do anything special (other than you would do ordinarily) to prepare students for the Alabama Direct Writing Assessment? YES NO

If YES, please describe.

_____

_____

_____

_____

10. If the Alabama Direct Writing Assessment were no longer required, do you believe writing instruction would change? YES NO

If YES, in what way(s)?

_____

_____

_____

_____

11. Do you have any recommendations for improving the Alabama Direct Writing Assessment? YES NO

If YES, how?

_____

_____

_____

_____

12. The following space is for the interviewer to record any additional observations or impressions that might be important in the evaluation.

_____

_____

_____

_____

Appendix B
Alabama Direct Writing Assessment Survey

Conducted by the Center for Educational Accountability
The University of Alabama at Birmingham
for the
Alabama State Department of Education

May 1996

School Code:_____ (for analysis purposes only)

Instructions: *Please circle your choice(s) for each question.*

1. Do you teach writing skills as part of your teaching duties? YES NO

2. At what level is your Alabama teaching certification? B A AA

3. Do you teach students writing at grade 5 YES NO
   or grade 7? YES NO

4. How many years have you taught writing?
   1 year or less
   2-5 years
   6-15 years
   more than 15 years

5. Have you been directly involved in the development of a state direct writing assessment (e.g., advisory committee, content and bias review)? YES NO

   If YES, in what capacity?_____

6. Have you administered a direct writing assessment in your school? YES NO

7. Are you familiar with the state's direct writing assessment scoring rubric? YES NO

8. Have you used the Alabama scoring rubric in your teaching of writing (e.g., for planning purposes, for scoring of student writing)? YES NO

   If YES, please describe how you have used the rubric.

   _____

   _____

   _____

   _____

Please continue on the back side.

Using the following scale, please rate your agreement/disagreement for items 9-20 by circling your choice:

Strongly Agree				Strongly Disagree	Not Applicable/ Don't Know
1	2	3	4	5	NA/DK

9. I have made specific efforts to align my writing instruction with the state's writing assessments.	1	2	3	4	5	NA/DK
10. I like my state's efforts to improve instruction through their implementation of mandatory assessments in writing.	1	2	3	4	5	NA/DK
11. I have received support and encouragement from my school administrators related to the state mandated writing assessments.	1	2	3	4	5	NA/DK
12. Preparing my students to take mandated writing assessments has had a positive effect on student writing skills.	1	2	3	4	5	NA/DK
13. The state direct writing assessment is based on sound principles of writing.	1	2	3	4	5	NA/DK
14. Teaching my students to use the scoring rubrics for the state mandated writing assessments has increased student confidence in their writing	1	2	3	4	5	NA/DK
15. I believe that the state mandated assessment results are an accurate reflection of student writing performance.	1	2	3	4	5	NA/DK
16. I have improved instruction and assessment in my English/language arts classes as a result of my state's implementation of the direct writing assessments.	1	2	3	4	5	NA/DK
17. I believe that the state mandatory writing assessments have had a negative impact by narrowing the variety of writing taught and practiced in the classroom.	1	2	3	4	5	NA/DK
18. As a result of the statewide direct writing assessments, my teaching of writing has improved.	1	2	3	4	5	NA/DK
19. As a result of the statewide direct writing assessments, I believe that the curriculum taught in English/language arts is becoming more uniform statewide.	1	2	3	4	5	NA/DK
20. I believe the state mandatory writing assessments have encouraged teachers of writing toward a common curriculum.	1	2	3	4	5	NA/DK

Thank you for your participation. Please return the completed survey directly to the Center for Educational Accountability, UAB School of Education, 901 13th Street, South, Birmingham, AL 35294-1250 in the prepaid envelope or fax it to 205-975-5389.

Ordering Information

1. Please complete this order form and mail with check or purchase order to SERVE, 345 South Magnolia Drive, Suite D-23, Tallahassee, Florida 32301. Make check or purchase order out to SERVE (Federal ID 59-6001-468).

2. Discounts are available for most SERVE products when you purchase units of fifty or more. Please call 800-352-6001 for discount information.

3. If you are requesting an invoice, your order must include a purchase order number.

4. We ship by U.S. Mail and United Parcel Service. Please calculate your shipping charges from the table below. Shipping charges will be higher for special orders and shipping outside the continental U.S. Please allow two weeks for delivery from the time we receive the order in our office. If you require special shipping arrangements, let us know. In most cases, we can accommodate your needs. Publication prices are subject to change.

5. For more information regarding SERVE's products and ordering procedures, please call 800-352-6001.

Shipping and Handling Charges

Up to $30.00	$ 2.50
$30.01 to $ 60.00	$ 5.00
$60.01 to $100.00	$ 7.50
$100.01 to $150.00	$10.00
$150.01 to $200.00	$12.50
$200.01 and above	call for price

Publications Listing

DESCRIPTION	ITEM #	PRICE
Action Research: Perspectives from Teachers' Classrooms	MSARP	$12.00
Appreciating Differences: Teaching and Learning in a Culturally Diverse Classroom	**HTADI**	**$10.00**
Assessment in Early Childhood Education: Status of the Issue	ECESI	$ 1.00
Children Exposed to Drugs: Meeting Their Needs	**HTSEC**	**$10.00**
Children Exposed to Drugs: What Policymakers Can Do	PBCED	$ 1.00
Comprehensive School Improvement	**HTCSI**	**$ 8.00**
Continuity in Early Childhood Education: A Framework for Home, School, and Community Linkages	ECECE	$12.00
Designing Teacher Evaluation Systems that Support Professional Growth	**RDTES**	**$ 8.00**
Does Class Size Make a Difference?	RDCSD	$ 4.00
Ed Talk: What We Know About Mathematics Teaching and Learning	**EDMAT**	**$ 7.00**
Ed Talk: What We Know About Reading Teaching and Learning	EDRTL	$ 7.00
Ed Talk: What We Know About Science Teaching and Learning	**EDSCI**	**$ 7.00**
Evaluation of the Alabama Direct Assessment of Writing Program	RDADE	$ 4.00

BEST COPY AVAILABLE

Families and Schools: An Essential Partnership	SSFSP	$ 6.00
Future Plans Planning Guide	FPPLG	$ 8.00
Going to Scale with TQM: The Pinellas County Schools' Journey Toward Quality	SSPCS	$12.00
How to Assess Student Performance in Science: Going Beyond Multiple-Choice Tests	RDSPS	$10.00
How to Improve Schools Now: Accessing SERVE's Programs, Products, and Publications	PRHIS	FREE
Improving Student Motivation: A Guide for Teachers and School Improvement Leaders	RDISM	$12.00
Interagency Collaboration: Improving the Delivery of Services to Children and Families	HTICO	$12.00
Issues to Consider in Moving Beyond a Minimal Competency High School Graduation Test	RDMCT	$ 4.00
Learning By Serving: 2,000 Ideas for Service Learning Programs	HTLBS	$ 8.00
A New Framework for School Accountability Systems	RDFRA	$ 3.00
Overcoming Barriers to School Reform in the Southeast	RDBAR	$ 3.00
Planning for School Improvement: A Report on a Comprehensive Planning Process	SRPSI	$ 1.00
Promising Service-Learning Programs	SSPSL	$ 1.00
Reducing School Violence: Building a Framework for School Safety	HTRSV	$ 8.00
Regional SERVE Forum on School Improvement: Proceedings (1996)	SIPROC	$ 8.00
Reengineering High Schools for Student Success	HTRHS	$ 8.00
Resources for School Improvement: How to Improve Schools Now	HTRSI	$10.00
Safe Schools: What the Southeast is Doing	PBSSC	$ 1.00
School Board Member Training in the Southeast	RDBMT	$ 4.00
Schools for the 21st Century: New Roles for Teachers and Principals	HTSTC	$ 8.00
Science in the Elementary Classroom: Portraits of Action Research	MSARE	$12.00
SERVE Catalog of Products and Publications	CATPP	FREE
SERVE-Line: Electronic Bulletin Board and On-Line Information System	SLEBB	FREE
South Pointe Elementary School (Year 1): A Public-Private Partnership	RDSP1	$ 1.00
South Pointe Elementary School (Year 2): A Public-Private Partnership	RDSP2	$ 1.00
South Pointe Elementary School (Year 3): A Public-Private Partnership	RDSP3	$ 1.00
Southern Crossroads: A Demographic Look at the Southeast	SRSCR	$ 3.00
Supporting Family Involvement in Early Childhood Education: A Guide for Business	SRSFI	$ 5.00
Technology Infrastructure in Schools *and* Plugging In: Choosing and Using Technology	PITI	$ 5.00
Together We Can: A Guide for Crafting a Profamily System of Education and Human Services	SRTWC	$ 8.00
Total Quality Management: Passing the Fad or "The Real Thing"? An Implementation Study	RDTQM	$ 5.00
Using Accountability as a Lever for Changing the Culture of Schools: Examining District Strategies	RDUAL	$ 8.00
Using Technology to Improve Teaching and Learning	HTTEC	$ 8.00
Youth Apprenticeship: A School-to-Work Transition Program	HTYAP	$ 8.00

BEST COPY AVAILABLE

Videotape Listing

Drug-Free Schools: A Generation of Hope (Running time: 27:00)	VTDFS	$19.95
Future Plans Videotape: Making the Most of Technology in the Classroom (Running time: 27:10) and Discussion Guide	FPPAK	$19.95
Passages: Providing Continuity from Preschool to School (Running time: 32:25)	VTPST	$19.95
School Improvement: Journey Toward Change (Running time: 23:00)	VTCSI	$19.95
Southern Crossroads: A Demographic Look at the Southeast (Running time: 22:00)	VTSCR	$19.95
Southern Solutions in Improving Mathematics and Science: General Audiences (Running time: 27:00)	VTMS3	$19.95
Southern Solutions in Improving Mathematics and Science: Policymakers (Running time: 60:00)	VTMS6	$19.95
Southern Solutions in Improving Mathematics and Science: Teachers/Practitioners (Running time: 1:24:00)	VTMS9	$19.95

Training and Seminars

For information on these training programs, please call 800-352-6001.	
Leadership for Collaboration: A Training Program	TRNLC
Providing a Safe and Healthy School Community	TRNSH
For information on these training programs, please call 800-545-7075.	
Comprehensive Crisis Management	TRNCC
Legal Principles Related to School Violence, Safety, Security, and Student Management Discipline	TRNLP

BEST COPY AVAILABLE

Order Form

Name: _____

Title: _____

Address: ❑ home ❑ work _____

City: _____ State: _____ Zip: _____

Phone: ❑ home ❑ work (_____) _____

Fax: ❑ home ❑ work (_____) _____

Quantity	Description	Item No.	Unit Price	Total

Subtotal	
Non-exempt Florida residents add 7% sales tax	
S & H	
Total	

Mail to:
SERVE
345 South Magnolia Drive
Suite D-23
Tallahassee, Florida 32301

Please photocopy this form for future orders.

NOTE: Ordering information and this form are located in the back of most SERVE publications.

Florida Tax Exemption =

BEST COPY AVAILABLE